

Folding of a fractal globule – the key for artificial molecular machines design

V.A. Avetisov¹, V.A. Ivanov², D.A. Meshkov¹, S.K. Nechaev^{3,4}

¹*N.N. Semenov Institute of Chemical Physics, RAS, 119991 Moscow, Russia*

²*M. V. Lomonosov Moscow State University, 119991 Moscow, Russia*

³*Université Paris-Sud/CNRS, LPTMS, UMR8626, Bât. 100, 91405 Orsay, France*

⁴*P.N. Lebedev Physical Institute, RAS, 119991 Moscow, Russia*

Folding and unfolding of a fractal (crumpled) polymer globule, which is recently proved to be a condensed structure of DNA in human chromosomes, is described as a cascade of equilibrium phase transitions in a hierarchical system. The fractal globule is characterized by a set of order parameters, encoding the crumples of different hierarchical levels. These order parameters have natural representation in terms of coordinates in the “ultrametric space”. Unfolding manifests itself in a sequential loss of stability of hierarchical levels with the temperature change, and is described by a master equation for a stochastic process with the jumps given by the Parisi-type matrix. Studying the relaxation properties of the elastic network of contacts in a fractal globule, we show that the dynamic properties of hierarchically folded polymer chains in globular phase are similar to those of molecular machines.

PACS numbers: 61.41.+e, 64.60.al, 64.60.aq, 87.15.Cc

A notion of molecular machine (MM) is usually attributed to a nanoscale structure able to convert perturbations of fast degrees of freedom into a slow motion along a specific path on a low-dimensional manifold – see [1–3]. Molecular machines possess various functions: they create or break chemical bonds, move mesoscopic objects, etc. Rich diversity of MMs is produced in a living world. Globular proteins formed by polypeptides of hundreds units look as simplest examples of MM. The ability of proteins to manipulate by solitary charges, or atoms against fluctuations, is the main requirement for reproduction of living systems at a molecular level [4, 5].

Is it possible to design molecular machines artificially? This challenging question addresses to the prospects of artificial algorithmic chemistry. In particular, it concerns the problem of overcoming the “complexity threshold” in prebiology, meaning that even a “primary” cell with minimal machinery of biological replication remains hopelessly complex for its own appearance beyond the world of MMs (see, for example, [6]). So, the question whether nonspecific polymers can spontaneously fold into MMs is still open. In this paper we propose the constructive way of creation one family of such structures.

In the definition of MMs we follow the work [3], where a MM is specified by a particular relaxation dynamics of its elastic network. Let a molecular structure be represented by a spatial distribution of a set of nodes $i = 1, \dots, N$ and undirected links between them, encoded in an adjacency matrix \mathbf{A} . The elements of \mathbf{A} are: $a_{ij} = 1$ for a link between i and j , and $a_{ij} = 0$ otherwise. In elastic networks, linked nodes are subjected the action of elastic forces that obey the Hooke’s law. In the over-damped limit, the velocity of each node is proportional to the sum of elastic forces applied to it, so the relaxation is:

$$\frac{d\mathbf{R}_i}{dt} = \sum_{j=1}^N a_{ij} \mathbf{u}_{ij} \left(|\mathbf{R}_i - \mathbf{R}_j| - |\mathbf{R}_i^{(0)} - \mathbf{R}_j^{(0)}| \right) \quad (1)$$

where $\mathbf{R}_i \equiv \mathbf{R}_i(t)$ is the current position of the i -th node and $\mathbf{R}_i^{(0)}$ is its equilibrium position and $\mathbf{u}_{ij} = \frac{\mathbf{R}_i - \mathbf{R}_j}{|\mathbf{R}_i - \mathbf{R}_j|}$. For small deviations from equilibrium, $|\mathbf{r}_i| = |\mathbf{R}_i - \mathbf{R}_i^{(0)}|$, nonlinear equations (1) can be approximated by linear ones

$$\frac{d\mathbf{r}_i}{dt} = - \sum_j \mathbf{\Lambda}_{ij} \mathbf{r}_j \quad (2)$$

where the strain tensors, $\mathbf{\Lambda}_{ij}$, are the building blocks of the $3N \times 3N$ linearized matrix $\mathbf{\Lambda}$. The relaxation of a system under small perturbations is a sum of independent normal modes, $\mathbf{r}_i(t) = \sum_{k=1}^{3N} \mathbf{r}_i(0) \mathbf{e}_k \exp(-\lambda_k t)$, where $\lambda_k > 0$ and \mathbf{e}_k are the eigenvalues and eigenvectors of the matrix $\mathbf{\Lambda}$. The smallest nonzero eigenvalues describing the relaxation of the system along the slowest degrees of freedom, are crucial for the very definition of MMs.

The distinguished features of MMs became transparent from the study of protein elastic network relaxation [3]. Being perturbed, a protein molecule quickly reaches a low-dimensional attracting manifold spanned by slowest degrees of freedom, and then slowly relaxes to the equilibrium along a particular path in this manifold. The existence of a low-dimensional manifold with a large basin of attraction and wide spectral gap separating slowest (soft) and fast (rigid) modes characterize the MMs.

The polymer globules with relaxation similar to MMs are the targets for our search. The globules obtained in the ordinary coil-to-globule phase transition have almost random elastic networks with many local metastable states and small basins of attraction. The soft and rigid relaxation modes of random networks are not well separated [3] which means that the randomly folded equilibrium polymer globules do not possess characteristics of MMs.

The search for candidates for a prototype of an artificial MM is based on the theoretical analysis of a protein

dynamics established by a variety of experimental data on CO–myoglobin binding [7], and on spectral diffusion in deeply frozen proteins [8]. These data point to the fact that protein fluctuational dynamics is similar to the *ultrametric* random walk in a wide temperature range from 4⁰K up to 300⁰K [9–11]. This result definitely signifies that proteins may have self–similar hierarchical order hidden in energy landscapes and conformational states (see also [12]).

A hierarchical organization, being translated to a network of contacts, means that a corresponding adjacency matrix has a block–hierarchical form similar to the Parisi matrix [13, 14]. Topological features of random block–hierarchical networks and their relevance to hierarchical systems of different nature, have been discussed in [17–19]. Recent work [15] has provided experimental evidence of the hierarchical structural organization of a DNA chain in a human genome in a form of a fractal globule (FG), which was predicted theoretically in [16] as a collapsed state of unknotted ring polymer. The apparent simplicity of a FG creation (the abrupt change of the solvent quality provokes a collapse transition in a long unentangled polymer ring which initially is in a coil state), has led to regard FG as a strong candidate for a molecular machine. Understanding the functioning of such macromolecular structures is impossible without developing an adequate language for the description of the process of creation and destruction of crumples on different hierarchical levels. Here we make first modest steps towards the construction of the equilibrium statistical theory of FG formation and melting.

The deep relation between phase transitions in a system of Ising spins with block–hierarchical interactions [20, 21] and critical phenomena in block–hierarchical networks, discussed in [19], allows us to suggest the new way of analytic description of folding and unfolding of the fractal globule. In the hierarchical Ising system, when temperature decreases, a hierarchy of phase transitions occurs producing a tree of hierarchically embedded sets of correlated spins. The theory of FG formation is based on the same ideas. It should be emphasized that the consideration presented below *is not a theory of coil–to–fractal globule* phase transition, but rather the *description of a FG formation and melting*, since we are interested in formation and destruction of hierarchical set of folds only, which happens *within* the globular phase of a polymer.

Consider a polymer chain in the $D = 3$ –dimensional space. The non–phantomness leads to two types of interactions: bulk interactions vanishing in infinitely thin chains, and topological interactions, preserved even for zero’s thickness chains. At $T > \theta$, a polymer of N segments, each of length ℓ , is in a coil phase. While at $T < \theta$, a macromolecule collapses into a weakly fluctuating drop–like globule of size $R_{gl} \sim \ell N^{1/3}$ ([22, 23]). In the globular phase formed by a linear chain *with open ends*, all subchains of length ℓs , where $s \lesssim N^{2/3}$, look as Gaussian coils due to the De Gennes theorem of screening

of volume interactions in the melt.

However, for linear *unknotted polymer rings* the situation is quite different. The topological constraints play the role of auxiliary repulsion and at $T < \theta$ there exists a certain critical length, ℓs_0 , (depending on temperature and bulk interaction), such that relatively small chain parts of the length ℓs_0 collapse. Define ℓs_0 –link segments as the block monomers, constituting the 1st–level crumples. The chain segments containing several consecutive 1st–level crumples collapse “in its own volume”, forming the 2nd–level crumples, they then collapse in the 3rd–level crumples, and so on. The hierarchical chain packing is completed when all ℓs_0 –link segments belong to a single crumple of a highest level. Such a packing means that a chain path is space–filling with the fractal dimension $D_f = D = 3$.

In a hierarchically folded macromolecule each s_0 –link unit can be characterized by a set of indices specifying to which particular 1st–level fold (embedded into particular 2nd–level fold, that, in turn, is embedded into particular 3rd–level fold, *etc*) this unit belongs. Such indexing is coded by a path on a Cayley tree – see the Fig.1. The boundary nodes (leaves) of this tree constitute a space of states for the units, and each subtree corresponds to a particular crumple. In case of regularly branching Cayley tree with the branching index $p + 1$ ($p = 2$ in the figure) and $\gamma = 1, 2, \dots$, enumerating the levels of hierarchy, each γ –level crumple consist of p^γ units. Hence, assuming that any two states are distinguished only at the scale $\sim p^\gamma$ of a maximal fold separating them, we end up with the *ultrametric* space of states. In such a description, the conformations of hierarchically folded chain are in bijection with the trajectories of a walker in an *ultrametric space*, which is the boundary of the tree. Thus, the chain conformation (hierarchy of folds) is fully characterized by the leaves (black dots) in the Fig.1a.

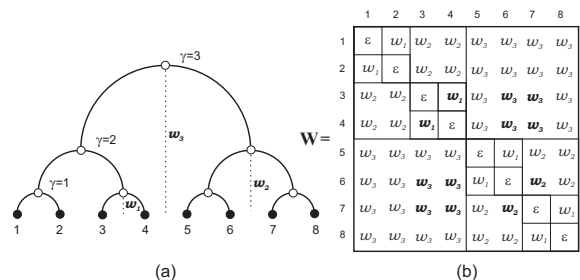


Figure 1: (a) Space of states in a fractal globule. Each chain monomer is characterized by its position on the boundary of the ultrametric space (black dots); (b) Block–hierarchical structure of transition matrix ($p = 2$).

Let $f_i(s)$ be the probability to find a walker in the state i at the boundary of the tree after s steps ($0 < s < N$), if at $s = 0$ it was at i_0 . Assuming the random walk to be a

homogeneous Markov chain, we have a master equation

$$\frac{\partial f_i(s)}{\partial s} = \sum_{j \neq i} w(i|j) f_j(s) - \left[\sum_{j \neq i} w(j|i) \right] f_i(s) \quad (3)$$

subject to the initial condition $f_i(s=0) = \delta_{i,i_0}$, where $w(i|j) = w(j|i)$ is the probability to jump between states i and j in one step. The ultrametric random walk is specified by the Parisi-like transition matrix \mathbf{W} of a block-hierarchical form, shown in the Fig.1b and obeying the strong triangle inequality (see, for example, [24]). For a random walk at the boundary of a regular p -branching Cayley tree, each block of the matrix \mathbf{W} of the hierarchical level γ consists of p^γ equal elements, that depend on γ only. For example, $w(3|4) = w_1$, $w(3|7) = w_3$, $w(5|8) = w_2$, etc – compare Fig.1(left) and Fig.1(right). The energy landscape of the ultrametric space is specified by a choice of transitions w_γ on the hierarchical level, γ . The landscape is supposed to be linear [9], i.e. $w_\gamma = p^{-(\alpha+1)\gamma}$, where $\alpha > 0$ is some parameter.

The conformations of the hierarchically folded chain of length $L = \ell N$ are parameterized by the N -step trajectories on the boundary of the Cayley tree. The transition probability $f_i(s=N)$ should be understood as a measure of all conformations of the hierarchically folded s -step chain, whose first unit is in an initial state i_0 and the last one, $s=N$, to be in the state i . The function $f_i(s)$ satisfies the Kolmogorov–Feller equation (3). It is known ([26]) that the symmetric transition operator in the right hand side of the equation (3) is diagonalizable by the p -adic Fourier transformation, and for unbounded ultrametric space its non-zero eigenvalues are $\lambda_\gamma = -p^{-\alpha\gamma}$.

Since all eigenvalues are negative, the hierarchically folded conformations given by (3) have a well defined ground state with a specific *set of eigenvalues, one for each fold of a given scale*. The chain conformations are subjected to the action of a hierarchy of strong constraints, and the temperature is supposed to be much lower than the critical temperature of the coil-globule phase transition.

The unfolding at some scale γ_{cr} means the destruction of hierarchical tree-like set of states above γ_{cr} . This can be easily achieved by introducing an auxiliary temperature-dependent repulsion between the units. The truncation of these interactions depends on the scale of crumples. To satisfy this requirement, the interaction potential should be again of a block-hierarchical form \mathbf{U} with elements depending on the ultrametric distances between the units only. Thus, we can generalize (3) to:

$$\frac{\partial f_i(s)}{\partial s} = \sum_{j \neq i} w(i|j) f_j(s) - \left[\sum_{j \neq i} w(i|j) \right] f_i(s) + \tau \mathbf{U} f_i(s) \quad (4)$$

where $\tau = \frac{T-\theta}{\theta}$ and $\mathbf{U} f_i(s) = \sum_{j \neq i} w(i|j) f_j(s)$

Again, using the p -adic Fourier technique one can diagonalize the transition operator in the right hand side

of (4). The nonzero eigenvalues are:

$$\lambda_\gamma = -p^{-\alpha\gamma} + \tau \left(-p^{-\alpha\gamma} + \frac{1-p^{-1}}{p^\alpha - p^{-1}} p^{-\alpha} \right) \quad (5)$$

The solutions of the equations $\lambda_\gamma(T) = 0$ define a sequence of critical temperatures $T_1 < T_2 < \dots < T_{max}$ at which the hierarchically folded conformations lose the stability. At T_1 the largest fold, i.e. the highest level γ_{max} becomes unstable and the fractal globule unfolds into a chain with smaller crumples. At $T_2 > T_1$, these crumples of level $\gamma = \gamma_{max} - 1$ unfold, etc, until T_{max} is reached, at which the crumple $\gamma = 1$ becomes unstable and the chain unfolds completely. If the temperature is decreasing, the hierarchical folding of a polymer chain is characterized by a sequence of critical temperatures, $T_{max} > \dots > T_2 > T_1$, at which the sequence of folds of hierarchical levels $\gamma = 1, 2, \dots, \gamma_{max}$ is formed.

We have used the Monte–Carlo method in the 3D continuous space [27] to simulate hierarchical folding of a polymer chain. A polymer of 202 monomers, among which we have chosen 32 specific units, is simulated in a large enough box without periodical boundary conditions. The chain is described by a “beads-on-the springs” model with a standard set of interaction potentials (see [27] for details), accounting for rigidity and volume interactions. In addition, a hierarchical potential U_γ operating between the specific units is introduced. It ensures that at each hierarchical level γ , the formation of clusters of specific units becomes possible only if two clusters of the previous level $\gamma - 1$ are close to each other in space, and vice versa: a cluster of a level $\gamma - 1$ cannot disappear if it is a part of an existing cluster of the level γ . A spatial distance between two clusters of the level γ is determined as the smallest distance between nodes of these clusters. We have used a short-range square-well potential with the depth of the well decreasing upon increasing the hierarchical level γ , but we have checked that the particular choice of this hierarchical potential is not crucial. However, the rules of hierarchical formation of clusters are very important. These rules imitate the hierarchical constraints imposed on the chain conformations under the polymer crumpling. The chain conformations are evolved via Metropolis algorithm. Equilibrating them, we construct an elastic network: specific units become the nodes of this network connected by *new* springs if the distances between them are less than three diameters of a unit; all other units are erased. A typical conformation of hierarchically collapsed polymer chain and its elastic network are shown in Fig.2a and Fig.2b respectively. The benchmarks A, B and C are used to visualize the relaxation trajectories of the elastic network (see [3] for details).

To characterize our system we have introduced the hierarchy of order parameters, $q_\gamma = \frac{Z(\gamma)}{Z_{max}(\gamma)}$ ($\gamma = 1, 2, \dots, \gamma_{max}$), where $Z(\gamma)$ is the number of actually formed clusters on a level γ averaged over a set of hierarchically folded structures, and $Z_{max}(\gamma)$ is the maximal

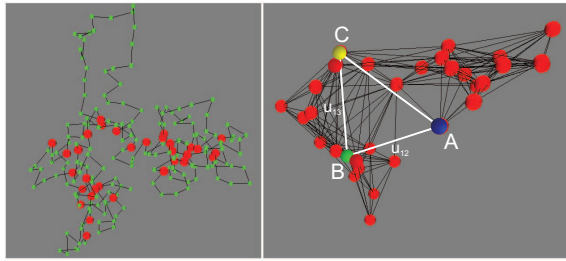


Figure 2: Left: Typical shape of hierarchically folded globule. Representative nodes are shown in red, other monomers are green; Right: The elastic network; blue, green and yellow colors mark the nodes constituting the 3D coordinate system for the visualization of the network dynamics.

number of possible clusters on the level γ , $0 \leq q_\gamma \leq 1$ for any γ . In the Fig.3(left) the simulations for the series of order parameters q_1, \dots, q_5 as the function of the inverse temperature, β , are shown. By increasing β we see the sequence of transitions: first, the smallest crumple is formed (black squares), then the 2nd crumple appears (red circles), *etc.* The typical eigenvalue spectra of elastic networks of hierarchically collapsed chains is shown in the Fig.3(right). Note a large spectral gap separating the slowest mode from the rest of spectrum ($\lambda_2/\lambda_1 \approx 7$, $\lambda_2/\lambda_3 \approx 1$).

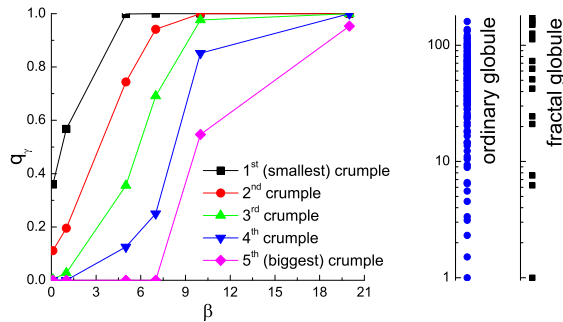


Figure 3: Left: Numerical observation of the hierarchy of transitions under the fractal globule formation: $\beta \sim (T^{-1} - \theta^{-1})$; Right: Eigenvalue spectra of the network of hierarchically folded chain and ordinary globule normalized to the lowest nonzero eigenvalue.

In the Fig.4 a 3D view of relaxation trajectories of the hierarchically folded elastic network (1) is shown. The relaxation dynamics is such that the hierarchically folded globule quickly relaxes to a low-dimensional (one-dimensional, in our demonstration) manifold with a large attracting basin and then slowly moves along it to the equilibrium. Such a behavior is very typical for molecular machines. It should be pointed out that the elastic network of the fractal globule is highly anisotropic: along

the δu_{23} -axis it is 10^3 more rigid than along δu_{12} - and δu_{13} -axes. In the ordinary globule such anisotropy is absent and its elastic network does not have any low-dimensional attracting manifold for phase trajectories. To represent the slowest degrees of freedom, we chose in the network a triangle ABC (marked by blue, green, and yellow nodes in the Fig.2) such that the changes of the distances along the sides AB ($\delta u_{12}/u_{12}^{(0)}$) and BC ($\delta u_{13}/u_{13}^{(0)}$) correspond to the slowest relaxation modes. The particular selection of benchmarks A, B and C is as follows. Computing the spectrum of elasticity matrix, we extract the eigenvectors, \mathbf{e}_1 and \mathbf{e}_2 , corresponding to slowest relaxation modes. Then the points A, B and C are chosen such that the projections of the average change of the distances AB onto \mathbf{e}_1 and BC onto \mathbf{e}_2 are maximal.

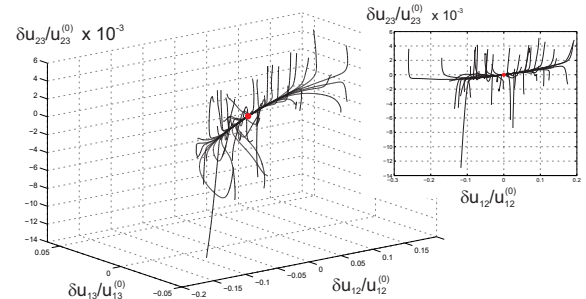


Figure 4: A set of relaxation trajectories of a hierarchically folded globule. Insert: Projection of relaxation trajectories on the plane $\delta u_{23}/u_{23}^{(0)}$ (fast mode) vs $\delta u_{12}/u_{12}^{(0)}$ (the slowest mode).

The structure considered here is typical for a wide family of hierarchical globules. The diversity concerns mainly the number of principal soft modes, manifested in the dimensionality of the attracting manifold. Typically, these manifolds are two- or three-dimensional. This fact might be important to get functional variability without altering the structural archetype. The considered in this work sample of FG does not realize yet any specific function except the conversion of thermal random perturbations of fast degrees of freedom into slow motion along a well-defined path. This capability allows, in principle, to manipulate by single atoms, or molecules in a controlled manner. The hierarchically folded domain might play in a molecular machine the same constructive role as an engine in a mechanical device.

The authors are grateful to A.S. Mikhailov and L. Mirny for encouraging discussions. This work was partially supported by the grant ANR-2011-BS04-013-01 WALKMAT and by MIT-France Seed Fund. The numerical part has been performed on the Supercomputing Center of MSU [28].

- 653–669 (1995)
- [3] Yu. Togashi and A.S. Mikhailov, Proc. Nat. Acad. Sci. USA **104** 8697–8702 (2007)
 - [4] M. Eigen, J. Mc-Cascill and P.J. Schuster, J. Phys. Chem. **92** 6881–6887 (1988)
 - [5] V.A. Avetisov and V.I. Goldanskii, Proc. Nat. Acad. Sci. USA. **93** 11435–11442 (1996)
 - [6] E.V. Koonin, Nat. Rev. Microbiol. **1** 127–136 (2003)
 - [7] P.J. Steinbach, A. Ansary, J. Berendzen, D. Braunstein, K. Chu, et al Biochemistry, **30** 3988–4001 (1991)
 - [8] V.V. Ponkratov, J. Friedrich, J.M. Vanderkooi, A.L. Burin and Yu.A. Berlin, J. Low. Temp. Phys. **3** 289–317 (2006).
 - [9] V. A. Avetisov, A. Kh. Bikulov and S. V. Kozyrev, J. Phys. A: Math. Gen. **32** 8785–8791 (1999)
 - [10] V. A. Avetisov, A. Kh. Bikulov, S. V. Kozyrev and V. A. Osipov, J. Phys. A: Math. Gen. **35** 177–189 (2002)
 - [11] V. A. Avetisov and A. Kh. Bikulov, Biophys. Rev. Lett. **3** 387–396 (2008)
 - [12] H. Frauenfelder, Nature Struct. Biol. **2** 821–823 (1995)
 - [13] G. Parisi, Phys. Rev. Lett. **43** 1754–1756 (1979)
 - [14] M. Mezard, G. Parisi and M. Virasoro, *Spin glass theory and beyond* (World Scientific: Singapore, 1987)
 - [15] E. Lieberman-Aiden, N. L. van Berkum, L. Williams, M. Imakaev, T. Ragoczy, A. Telling, I. Amit, B. R. Lajoie, P. J. Sabo, M. O. Dorschner, et al., Science **326** 289–293 (2009)
 - [16] A.Yu. Grosberg, S.K. Nechaev and E.I. Shakhnovich, J. Phys. (Paris) **49** 2095–2100 (1988)
 - [17] V.A. Avetisov, A.V. Chertovich, S.K. Nechaev and O.A. Vasilyev, J. Stat. Mech. **07** 07008–07023 (2009)
 - [18] V.A. Avetisov, A.Kh. Bikulov, O.A. Vasilyev, S.K. Nechaev and A.V. Chertovich, Journal of Experimental and Theoretical Physics **109**(3), 485–504 (2009)
 - [19] V.A. Avetisov, S.K. Nechaev and A.V. Shkarin, Physica A, **389**, 5895–5902 (2010)
 - [20] F. J. Dyson, Comm. Math. Phys. **12** 91–107 (1962)
 - [21] G.A. Baker Jr., Phys. Rev. B **5** 2622 (1972)
 - [22] I.M. Lifshitz, A.Yu. Grosberg and A.R. Khokhlov, Some problems of the statistical physics of polymer chains with volume interaction, Rev. Mod. Phys. **50** 683–713 (1978)
 - [23] A. Yu. Grosberg and A. R. Khokhlov, *Statistical Physics of Macromolecules* (AIP Press: New York, 1994)
 - [24] A. Ogielsky and D. Stein, Phys. Rev. Lett. **55** 1634–1637 (1985)
 - [25] Parisi G. and Sourlas N., European Phys. J., **B 14**, 535–542 (2000)
 - [26] V.S. Vladimirov, I.V. Volovich and Y.I. Zelenov *p-Adic Analysis and Mathematical Physics* (Singapore: World Scientific, 1994)
 - [27] I. Gerroff, A. Milchev, K. Binder and W. Paul, J. Chem. Phys. **98** 6526–6539 (1993)
 - [28] V.I. Voevodin, S.A. Zhumatiy, S.I. Sobolev, A.S. Antonov, P.A. Bryzgalov, et al, Open Systems J., Moscow: Open Systems Publ. No. 7 (2012) [in Russian]